RESEARCH

Changes in Genetic Diversity of U.S. Flue-Cured Tobacco Germplasm over Seven Decades of Cultivar Development

H.S. Moon, J.S. Nicholson, A. Heineman, K. Lion, R. van der Hoeven, A.J. Hayes, and R.S. Lewis*

ABSTRACT

Plant breeding methodologies have been applied to flue-cured tobacco (*Nicotiana tabacum* L.) for approximately seven decades. As has been observed in several other crops, stringent quality requirements have resulted in use of conservative breeding strategies in the development of new cultivars. The impact of breeding practices on genetic diversity within U.S. flue-cured tobacco germplasm has not been investigated. In this study, we genotyped 117 tobacco cultivars from eight sequential time periods with 71 microsatellite primer pairs. A total of 294 alleles were scored. Only a fraction (48%) of alleles present in the initial germplasm pool was represented in cultivars released during the 1990s and 2000s. Only 13 and 18 alleles were detected in the 1990s and 2000s, respectively, which were undetected in the initial gene pool. The overall trend was one of gradual reduction in allelic counts at microsatellite loci, indicating a reduction in diversity over time at the gene level. Average genetic similarity was highest among cultivars of the 1990s and 2000s, reflecting a reduction in genetic diversity at the population level. This observed narrowing of the U.S. flue-cured tobacco germplasm base in combination with low rates of genetic gain for yield in the last 20 years may point to a need for diversification of parental materials used in future breeding crosses. Reported genetic relationships among the group of genotyped cultivars may be valuable for future strategic germplasm choices.

H.S. Moon, and R.S. Lewis, Dep. of Crop Science, North Carolina State Univ., Raleigh, NC 27695; J.S. Nicholson, USDA-APHIS Plant Protection and Quarantine, Raleigh, NC 27606; A. Heineman, and K. Lion, Lancaster Laboratories, Richmond, VA, 23234; R. van der Hoeven, A.J. Hayes, Philip Morris USA, Richmond, VA 23261. Received 9 May 2008. *Corresponding author (ramsey_lewis@ncsu.edu).

Abbreviations: AFLP, amplified fragment length polymorphism; RAPD, random amplified polymorphic DNA; PCO, principle coordinate analysis; PIC, polymorphic information content; SSR, simple sequence repeat; UPGMA, Unweighted Pair Group Method with Arithmetic Averaging.

Tobacco (*Nicotiana tabacum* L.) is one of the most economically important non-food crops cultivated worldwide. Application of scientific plant breeding methods to tobacco began in the early part of the 20th century (Shamel and Cobey, 1907; East and Jones, 1921, Garner et al., 1936; Clayton, 1958). Although substantial increases in yield and disease resistance have been achieved for the flue-cured class of tobacco since the 1940s (Wernsman and Rufty, 1987), stringent industry requirements for quality attributes (Bowman, 1996) have led to conservative breeding strategies. Genetic characteristics of various tobacco market classes appear to be unique and limit the amount of germplasm from one class that can be tolerated in another without adversely affecting yield or quality (Wernsman and Rufty, 1987). Breeding crosses are typically only made between elite materials from within the flue-cured market class in a practice termed "advanced cycle pedigree breeding" by Bernardo (2002). Novel genetic variation from non-flue-cured *N. tabacum* germplasm has only been incorporated in small doses in efforts to transfer relatively simply

Published in Crop Sci. 49:498–508 (2009).
doi: 10.2135/cropsci2008.05.0253

677 S. Segoe Rd., Madison, WI 53711 USA

inherited traits, such as genetic resistance to black shank (*Phytophthora parasitica* Dastur var. *nicotianae* (Breda de Haan) Tucker) and bacterial wilt (*Ralstonia solanacearum* E.R. Smith) (Bullock, 1943; Smith et al., 1945). Wild *Nicotiana* relatives have only been used as sources of simply-inherited disease-resistance genes.

Sustained genetic improvement of crop plants depends on availability of genetically variable populations in which selection can be conducted. The restricted nature of the U.S. flue-cured tobacco germplasm base has been suggested by pedigree analysis. Murphy et al. (1987) estimated the average coefficient of parentage among a set of 131 historically important U.S. flue-cured tobacco cultivars to be 0.41. This is considerably higher than similar determinations made for other self-pollinated cultivated crop species (Lewis and Nicholson, 2007). The extent to which modern breeding practices may have narrowed the immediate flue-cured tobacco germplasm pool over time has not been investigated. Such analyses can be important for understanding the impact of plant breeding on crop diversity and for devising strategies for future genetic improvement.

DNA markers have been useful for examining genetic diversity within crop species (Donini et al., 2000; Lu and Bernardo, 2001; Le Clerc et al., 2005; Fu et al., 2006). Past molecular marker research in tobacco has mostly focused on the use of randomly amplified polymorphic DNA (RAPD) or amplified fragment length polymorphism (AFLP) markers. These marker systems reveal relatively low levels of polymorphism among *N. tabacum* materials, however (Rossi et al., 2001; Ren and Timko 2001; Nishi et al., 2003), making them most useful for detecting chromatin introgressed from wild relatives (Bai et al., 1995; Yi et al., 1998; Johnson et al., 2002; Lewis 2005; Milla et al., 2005, Lewis et al., 2005; Moon and Nicholson 2007; Lewis et al., 2007). Large-scale sequencing efforts for tobacco, however, have led to the development of microsatellite, also referred to as simple sequence repeat (SSR), markers that have increased observed rates of polymorphisms in tobacco and that have been useful for genetic mapping in this species (Bindler et al., 2007).

For this research, a set of 117 flue-cured tobacco cultivars were genotyped with 71 microsatellite primer pairs described by Bindler et al. (2007). The first objective was to use this genotypic information to investigate genetic relationships among these 117 cultivars. The second goal was to determine the fraction of microsatellite alleles found in 37 representatives of the original germplasm pool of "farmer varieties" that are now present in modern cultivars. The third objective was to assess the degree to which application of scientific plant breeding methods has affected genetic diversity at the gene and population levels over seven decades of flue-cured tobacco cultivar improvement.

MATERIALS AND METHODS

Plant Materials

A total of 117 flue-cured tobacco cultivars divided into eight groups were selected for this research (Table 1). The first group consisted of a set of 37 flue-cured farmer varieties that were grown in tobacco growing regions of the U.S. in the early part of the 20th century (Garner et al., 1936; Chaplin et al., 1962). Almost all flue-cured tobacco germplasm is presumed to have originated from this gene pool. The remaining seven groups included cultivars released during seven decades: 1940s, 1950s 1960s, 1970s, 1980s, 1990s, and 2000s. Each of these decades was represented by 10 to 16 cultivars. Attempts were made to include all cultivars that occupied significant acreage during the given periods. Seeds of historical cultivars were obtained from the U.S. *Nicotiana* Germplasm Collection (Lewis and Nicholson, 2007). Seeds of recently released cultivars were obtained from private seed companies or public universities. Information related to each cultivar (year of release, etc.) was obtained from published tobacco extension documentation or historical North Carolina Official Variety Test information.

DNA Extraction

Genomic DNA was isolated using DNeasy plant DNA isolation reagents (Qiagen, Valencia, CA). Approximately 5 g of fresh leaf material (bulk of tissue from three greenhouse-grown plants) were ground for each cultivar using liquid nitrogen. For each sample, 10 mL of AP1 reagent (65°C) and 20 μL RNase A was added followed by vigorous shaking. Samples were then incubated for 10 min in a 65°C water bath and subsequently placed in ice for 10 min. After centrifugation for 5 min at 4000 rpm at room temperature, supernatants were transferred to fresh 50 mL conical tubes with 10 mL of 25:24:1 (v/v/v) phenol:chloroform:isoamyl alcohol (pH 8.0). After shaking for 15 min, tubes were centrifuged at 3000 rpm for 10 min at room temperature. Aqueous layers were transferred to fresh 50 mL conical tubes and 12 mL of 100% isopropanol was added. After centrifugation and removal of supernatants, DNA pellets were dried and then resuspended in TE buffer (pH 7.4).

Microsatellite Detection

All 117 cultivars were genotyped using 71 microsatellite primer pairs amplifying bands at loci positioned on 23 of 24 linkage groups (Bindler et al., 2007). A multiplex system of genotyping was used where two to six primer pairs were simultaneously analyzed using different fluorescent tags. PCR reactions were performed in 25 μL final volumes containing 25 to 50 ng of template DNA, 12.5 μL 2X Qiagen Multiplex PCR master mix (Qiagen), 2.5 μL fluorescently labeled primer mix (0.2 uM per primer), 1 μL 100% DMSO, and 8 μL H_2O (DNase/RNase free). Reaction mixes were overlaid with 10 μL of light mineral oil to reduce evaporation. Reactions were initiated with a 15 min incubation period at 95°C followed by 34 cycles of 1 min at 94°C, 2 min at 60°C, and 1 min at 72°C. The final reaction step was 60°C for 30 min.

Reaction products were diluted 1:50 (PCR mixture: water), and 2 μL of diluted products were then combined with 9.75 μL of HiDi Formamide (Applied Biosystems, Foster City, CA) and 0.25 μL of GS500LIZ (Applied Biosystems), or 0.25 μL

Table 1. 117 North American flue-cured tobacco cultivars chosen for study with their year of release and plant introduction number.

Cultivar	Code	Release year	PI no./source	Cultivar	Code	Release year	PI no./source
Big Gem	1	Farmer Variety	PI 552343	Virginia 45	60	1955	PI 557003
Burch's Special	2	Farmer Variety	PI 552303	Coker 187	61	1956	PI 552391
Cabbage	3	Farmer Variety	PI 552304	Coker 187-Hicks	62	1957	PI 552392
Delcrest	4	Farmer Variety	PI 552306	Coker 316	63	1959	PI 552394
Faucette Special	5	Farmer Variety	PI 552309	SC 58	64	1959	PI 552344
Griffin Special	6	Farmer Variety	PI 552311	McNair 10	65	1960	PI 552422
Harrison Special	7	Farmer Variety	PI 552312	NC 95	66	1961	PI 552380
Jamaica Wrapper	8	Farmer Variety	PI 552316	Speight G-10	67	1961	PI 551317
Dukane	9	Farmer Variety	PI 552308	Coker 319	68	1962	PI 552426
Yellow Mammoth	10	Farmer Variety	PI 552338	McNair 20	69	1962	PI 552429
Yellow Pryor	11	Farmer Variety	PI 552339	McNair 30	70	1962	PI 552430
Silver Dollar	12	Farmer Variety	PI 552332	Coker 298	71	1964	PI 552445
Southern Beauty	13	Farmer Variety	PI 552333	NC 2326	72	1964	PI 552453
Little Sweet Orinoco	14	Farmer Variety	PI 552376	Va. 115	73	1964	PI 552458
Lizard Tail Orinoco	15	Farmer Variety	PI 552377	Coker 258	74	1966	PI 552461
Lemon Bright	16	Farmer Variety	PI 552317	Speight G-28	75	1969	PI 551318
Adcock x Harrison Pryor	17	Farmer Variety	PI 552296	McNair 135	76	1970	PI 551304
Adcock x Pinckney Arthur	18	Farmer Variety	PI 552297	Coker 347	77	1971	PI 552462
D.H. Currin	19	Farmer Variety	PI 552307	SC 72	78	1972	PI 551316
Harrison Pryor	20	Farmer Variety	PI 552313	McNair 944	79	1973	PI 552494
Hickory Pryor	21	Farmer Variety	PI 552314	Speight G-15	80	1974	PI 552493
Pinckney Arthur	22	Farmer Variety	PI 552329	Coker 86	81	1976	PI 552489
Warne	23	Farmer Variety	PI 552335	Coker 48	82	1977	PI 552495
Virginia Bright Leaf	24	Farmer Variety	PI 552385	McNair 373	83	1979	PI 552373
Paris Wrapper	25	Farmer Variety	PI 552308	NC 82	84	1979	PI 551311
Robertson	26	Farmer Variety	PI 552330	Speight G-70	85	1979	PI 552497
Silk Leaf	27	Farmer Variety	PI 552331	Coker 51	86	1980	PI 552503
Banana Leaf	28	Farmer Variety	PI 552298	Coker 176	87	1982	PI 551294
Bonanza	29	Farmer Variety	PI 552300	K326	88	1982	Gold Leaf Seed Co.
Bottom Special	30	Farmer Variety	PI 552301	K399	89	1982	Gold Leaf Seed Co.
Broad Leaf Orinoco	31	Farmer Variety	PI 552302	NC567	90	1983	PI 552714
Cash	32	Farmer Variety	PI 552305	K 394	91	1984	Gold Leaf Seed Co.
Gold Dollar	33	Farmer Variety	PI 552310	Speight G-80	92	1984	Speight Seed Farms
Hicks Broadleaf	34	Farmer Variety	PI 552397	NC27NF	93	1986	PI 551309
Jamaica	35	Farmer Variety	PI 552315	K340	94	1986	PI 552667
White Mammoth	36	Farmer Variety	PI 552336	Coker 371-Gold	95	1987	PI 552524
White Stem Orinoco	37	Farmer Variety	PI 552337	NC37NF	96	1988	PI 552712
401	38	1942	PI 552342	K149	97	1990	Gold Leaf Seed Co.
Virginia Gold	39	1947	PI 552334	K346	98	1990	Gold Leaf Seed Co.
Yellow Special A	40	1943	PI 552378	RG 81	99	1995	Rickard Seed
Golden Wilt	41	1949	PI 552393	NC55	100	1996	Gold Leaf Seed Co.
Golden Harvest	42	1948	PI 552399	NC71	101	1996	Gold Leaf Seed Co.
White Gold	43	1949	PI 552400	Oxford 207	102	1996	Gold Leaf Seed Co.
Ox. 1–181	44	1948	PI 552401	NC72	103	1997	Gold Leaf Seed Co.
Dixie Bright 27	45	1949	PI 552355	Speight 168	104	1997	Speight Seed Farms
Dixie Bright 101	46	1949	PI 552383	Speight NF3	105	1997	Speight Seed Farms
Dixie Bright 102	47	1949	PI 552384	NC606	106	1999	Raynor Seeds
Oxford 1	48	1942	PI 552320	NC297	107	2000	Gold Leaf Seed Co.
Oxford 2	49	1942	PI 552321	GL 737	108	2000	Gold Leaf Seed Co.
Oxford 3	50	1942	PI 552322	Speight H20	109	2000	Speight Seed Farms
Oxford 26	51	1945	PI 552323	GL973	110	2001	Gold Leaf Seed Co.
Vamorr 48	52	1948	PI 552767	NC810	111	2001	Cross Creek Seeds
Vesta 30	53	1940s	PI 552769	Speight 210	112	2001	Speight Seed Farms
Golden Cure	54	1950	PI 552390	CU748	113	2003	Clemson University
Vesta 5	55	1952	PI 552396	NC291	114	2003	Cross Creek Seeds
Coker 139	56	1954	PI 552389	Speight 220	115	2003	Speight Seed Farms
Dixie Bright 28	57	1954	PI 552356	NC 471	116	2005	Raynor Seeds
Virginia 21	58	1955	PI 552398	CC 27	117	2005	Cross Creek Seeds
Dixie Bright 244	59	1955	PI 551300				

GS500ROX (Applied Biosystems). Samples were separated using a 36 cm capillary array on either an ABI 3100 or an ABI 3730 DNA sequencing system (Applied Biosystems). Amplicons were scored using the "Local Southern Method" and default analysis settings within GeneMapper (v. 3.5) software (Applied Biosystems). Final band sizes were standardized to an internal DNA control and on the basis of the ABI 3730 sequencing system.

Data Analysis

For each microsatellite primer pair, standard statistics were computed that included the total number of alleles, number of rare alleles (*frequency* < 0.05), and polymorphic information content (PIC) as described by Roussel et al. (2004).

Possible changes in genetic diversity over time were investigated using multiple approaches. To examine changes at the gene level, the total number of alleles produced per primer pair and the total number of alleles amplified by all primer pairs were calculated for each of the eight breeding periods. To statistically compare allelic count numbers for any two groups of cultivars, we applied the random permutation procedure of Fu et al. (2003). In this method, a PROC IML program (SAS Institute, Cary, NC) was first used to select a single allele. On the basis of the observed frequency of this allele in the group of 117 cultivars, the allele was allocated to the 117 cultivars without replacement regardless of year of release. This step was repeated for all other alleles identified in this investigation, followed by counting the number of alleles for the simulated cultivars from the given time periods. Differences in allelic counts between two groups of simulated cultivars were then calculated and compared with actual observed differences. These steps were then repeated a total of 10,000 times, and results were averaged over all permutations to produce the expected number of alleles and standard deviations for cultivars in each time period. The probability of observing the difference between actual and expected values in the absence of selection was given by the proportion of the 10,000 permutations where the difference in simulated allelic count was greater than the actual allelic difference.

Comparisons of average genetic similarities for each of the eight time periods were used to examine possible changes in genetic diversity at the population level. Genetic similarity values (S_{ij}) were calculated for all pairwise cultivar combinations within a given time period according to the method of Dice (1945), and mean S_{ij} values were subsequently calculated for each period. Standard errors for mean S_{ij} values and *t* tests for testing for significant differences between mean S_{ij} values for different time periods were calculated according to Leonard et al. (1999). The *t* tests applied here were designed for application to marker data where amplified PCR products are expected to be nearly independent, as is the case with RAPD or AFLP data. Leonard et al. (1999) indicate that, for microsatellite data (where different alleles amplified by a given primer pair are usually not independent), standard errors for mean S_{ij} are underestimated, although probably by a small amount. Thus, the *t* tests of Leonard et al. (1999) likely slightly overestimate the significance of differences between mean S_{ij}'s for our situation.

The genetic relationships among the 117 cultivars was also examined by first generating a genetic distance matrix using the shared allele coefficient (Chakraborty and Jin, 1993) by PowerMarker version 3.25 (Liu and Muse, 2005). The distance matrix was then imported into NTsys-PC version 2.2 (Rohlf, 2000) and a dendrogram was produced based on the matrix and the Unweighted Pair Group Method with Arithmetic Averaging (UPGMA) algorithm (Sneath and Sokal, 1973). Bootstrap support for dendrogram branches was conducted using 500 replicates by means of the software program Winboot (Yap and Nelson, 1996).

Cultivar associations were also investigated via principle co-ordinate (PCO) analysis. The genetic distance matrix based on the shared allele coefficient and generated by PowerMarker was imported into NTsys-PC version 2.2, and PCO analysis was performed using the Dcenter and Eigen functions. The first two axes from the analysis were plotted.

RESULTS

Microsatellite Polymorphism

Of the 71 microsatellite primer pairs used in this study, 69 were found to produce amplification products that were polymorphic amongst the 117 cultivars that were genotyped (Table 2). These 69 primer pairs amplified loci from 23 of the 24 linkage groups reported by Bindler et al. (2007), and a total of 294 alleles were scored. The number of alleles detected by these primers ranged from 2 to 12 (Table 2), with a mean of 4.26. PIC values for primer pairs that produced polymorphic bands ranged from 0.009 to 0.862 (Table 2). Of the 294 total observed alleles, 166 (56.5%) were considered to be rare (*frequency* < 0.05). One-hundred and twenty-four and 130 alleles that were detected in the initial gene pool were undetected for the 1990s and 2000s time periods, respectively. Only 13 and 18 alleles were detected in the 1990s and 2000s, respectively, which were undetected in the initial gene pool.

Changes in Allelic Diversity

Changes in the number of alleles detected per locus reflect changes in genetic diversity at the gene level. The overall trend for the total number of observed alleles was one of gradual reduction. The total allele count declined substantially from 215 for the initial gene pool to 169 for the 1940s. The only exception to the continuing trend of gradual reduction over time was an increase in the observed allele count from 125 for the 1960s to 135 for the 1970s. The 1990s and 2000s exhibited the lowest observed total allele counts (104 and 103 for each period, respectively). This represents approximately 48 and 61% of the total number of alleles present in the initial gene pool and the 1940s time period, respectively.

The permutation test of Fu et al. (2003) was applied to the genotypic data to reduce possible bias in comparing allelic count data for groups with different numbers of members. Results indicated significantly lower allelic counts for the 1960s, 1980s, 1990s, and 2000s relative to the total allelic count for the initial gene pool ($P < 0.003$) (Table 2). Because only a handful of cultivars from the

Table 2. Microsatellite primer pairs used and corresponding allelic diversity measures.

Primer pair	Linkage group†	Total allelic count	No. of rare alleles‡	PIC value	Allelic count for cultivars of various breeding periods								
					Initial gene pool (37)§	1940s (16)	1950s (11)	1960s (11)	1970s (10)	1980s (11)	1990s (10)	2000s (11)	Prob (E > O)¶
PT30259	1	3	2	0.065	3	1	1	1	1	2	1	1	0.172
PT30307	1	5	3	0.276	5	2	2	1	2	2	1	1	0.012
PT30424	1	3	2	0.034	2	1	1	1	1	2	1	1	0.402
PT30114	2	8	4	0.664	6	5	4	2	4	3	3	2	0.203
PT30242	2	3	2	0.057	3	1	1	1	1	1	1	1	0.174
PT30327	2	2	1	0.017	2	1	1	1	1	1	1	1	0.318
PT30375	2	4	1	0.588	4	4	3	3	3	3	1	2	0.040
PT30197	3b	3	1	0.372	2	3	3	2	3	2	2	1	0.555
PT30205	3b	6	5	0.171	2	2	5	2	3	3	1	2	0.938
PT30229	3b	2	1	0.025	1	2	2	1	1	1	1	1	0.880
PT30124	4	4	2	0.141	2	3	1	2	2	2	1	1	0.700
PT30471	5	3	2	0.034	2	2	1	1	1	1	1	1	0.461
PT30011	6	7	5	0.322	5	3	2	2	6	2	1	1	0.192
PT30087	6	3	1	0.334	3	3	2	2	3	2	2	2	0.554
PT30157	6	4	3	0.042	3	1	2	1	1	1	1	1	0.207
PT30449	6	6	2	0.522	3	2	3	3	3	4	4	3	0.933
PT30138	7	3	2	0.095	3	2	2	1	2	1	1	1	0.166
PT30202	7	6	5	0.165	3	4	3	2	1	1	1	2	0.784
PT30394	7	3	2	0.111	2	3	2	1	1	1	1	1	0.680
PT30164	8a	3	1	0.232	3	2	2	2	2	2	2	2	0.398
PT30388	8a	2	1	0.017	2	1	1	1	1	1	1	1	0.318
PT30361	8b	3	2	0.110	3	1	1	1	1	3	1	1	0.241
PT30044	9	7	5	0.309	6	4	2	2	4	2	3	2	0.121
PT30265	9	10	7	0.462	9	3	4	3	2	3	2	2	0.006
PT30416	9	4	2	0.152	3	2	2	2	1	1	1	1	0.261
PT30421	9	5	2	0.536	3	4	3	4	4	3	3	1	0.288
PT30077	10	3	2	0.068	3	2	2	1	1	1	1	1	0.168
PT30132	10	12	5	0.862	9	8	3	6	5	5	3	3	0.167
PT30250	10	4	3	0.104	1	2	3	1	2	2	1	2	0.987
PT30311	10	3	0	0.400	2	2	2	3	2	2	1	2	0.945
PT30380	10	4	2	0.328	2	3	3	3	3	2	2	2	0.904
PT30482	10	3	2	0.033	3	1	1	1	1	1	1	1	0.104
PT30350	11	6	3	0.253	5	3	3	2	1	2	2	2	0.212
PT30099	12	7	3	0.556	7	4	4	3	3	3	2	1	0.001
PT30324	12	4	3	0.119	2	1	1	2	1	1	1	2	0.912
PT30473	12	5	4	0.083	3	2	1	2	1	1	1	1	0.355
PT30137	13	3	1	0.271	3	3	3	2	2	2	2	2	0.597
PT30342	13	5	4	0.150	2	3	3	1	3	2	1	1	0.731
PT30214	14a	3	2	0.088	2	1	1	1	2	1	1	1	0.603
PT30403	14a	3	1	0.175	3	2	2	1	2	1	1	1	0.157
PT30159	14b	10	6	0.732	8	6	2	2	3	2	1	1	0.009
PT30188	16	3	1	0.131	2	2	2	2	2	2	2	1	0.537
PT30459	16	3	2	0.058	1	3	1	1	1	1	2	1	0.901
PT30053	17	3	1	0.308	2	3	2	2	2	2	2	2	0.890
PT30156	17	9	6	0.572	7	6	2	2	3	2	2	3	0.308
PT30339	17	4	3	0.050	2	2	1	1	1	1	1	2	0.886
PT30111	18	2	1	0.033	2	1	1	1	1	1	1	1	0.476
PT30163	18	2	0	0.139	2	2	2	2	2	2	2	1	0.355
PT30005	19	3	1	0.247	3	2	1	2	3	2	2	2	0.534
PT30165	19	4	3	0.042	2	2	1	1	1	1	1	2	0.880
PT30230	19	4	3	0.138	3	3	1	2	1	1	1	1	0.448
PT30248	19	4	3	0.066	2	2	2	1	2	1	1	1	0.624
PT30411	19	4	3	0.165	3	2	2	1	2	2	1	2	0.777
PT30150	20	2	1	0.033	2	1	1	1	1	2	1	1	0.475

Table 2. Continued.

Primer pair	Linkage group†	Total allelic count	No. of rare alleles‡	PIC value	Allelic count for cultivars of various breeding periods								
					Initial gene pool (37)§	1940s (16)	1950s (11)	1960s (11)	1970s (10)	1980s (11)	1990s (10)	2000s (11)	Prob (E > O)¶
PT30142	1 & 23	3	1	0.301	2	2	2	2	2	2	3	2	0.891
PT30235	21	4	2	0.377	3	3	2	2	2	2	1	1	0.159
PT30378	21	2	1	0.017	2	1	1	1	1	1	1	1	0.313
PT30028	22	8	5	0.639	6	4	4	4	3	3	3	3	0.439
PT30084	22	2	1	0.018	1	1	2	1	1	1	1	1	0.905
PT30095	22	3	2	0.095	2	2	1	1	2	1	1	1	0.618
PT30177	22	6	4	0.291	4	2	2	2	1	2	2	3	0.659
PT30364	22	3	2	0.110	2	2	2	1	1	1	1	1	0.659
PT30160	23	5	2	0.533	2	3	3	3	2	4	2	2	0.898
PT30186	23	4	1	0.316	4	3	3	3	1	1	1	1	0.025
PT30231	23	2	1	0.009	1	1	1	2	1	1	1	1	0.908
PT30257	23	5	1	0.541	4	2	2	3	4	3	3	3	0.689
PT30200	24	3	1	0.109	3	2	1	1	1	1	1	1	0.159
PT40005	24	8	6	0.441	4	4	2	3	3	3	3	2	0.606
PT40024	24	2	1	0.017	2	1	1	1	1	1	1	1	0.310
Observed total		294	166		215	169	138	125	135	125	104	103	
Observed lost#						88	104	110	105	111	124	130	
Observed new††						42	27	20	25	21	13	18	
Expected total‡‡					209	162	145	145	140	145	140	145	
Expected standard deviation §§					5.9	5.8	5.6	5.6	5.5	5.6	5.6	5.6	
Prob (E > O)¶						0.570	0.082	0.002	0.104	0.002	<0.001	<0.001	

†Linkage group conforms to that published by Bindler et al. (2007).

‡Rare alleles are those present with a frequency below 0.05.

§Number in parenthesis equals the total number of cultivars genotyped for each time period.

¶Equals the proportion of 10,000 random permutations conducted according to Fu et al. (2003) where the simulated difference in the number of alleles between the initial gene pool and the given time period was larger than the observed difference.

#Observed lost = the total number of alleles undetected in the cultivars of a specific time period relative to those present in the initial gene pool.

††Observed new = the total number of new alleles detected in the cultivars of a specific time period relative to those present in the initial gene pool.

‡‡Expected total = the total number of alleles expected to be detected in the cultivars of a specific time period as determined by the method of Fu et al. (2003).

§§Expected standard deviation = the standard deviation of the number of alleles expected to be detected in the cultivars of a specific time period as determined by the method of Fu et al. (2003).

initial gene pool are thought to have contributed to the development of modern-day cultivars (Murphy et al., 1987), genotypes for all members of the initial gene pool were eliminated from the analysis and comparisons for all groups were made relative to the 1940s time period. Similar results were again observed, where all periods had significantly reduced levels of allelic diversity relative to the 1940s ($P < 0.019$).

Trends for individual microsatellite loci were also examined. For 54 out of 69 microsatellite loci, the number of observed alleles for the 2000s was numerically lower than the total number for the initial gene pool (Table 2). There was an increase in the total number of alleles for only one primer pair (PT30250). Permutation tests indicated only six reductions on six linkage groups to be significant at the $P < 0.05$ level, however (Table 2).

Changes in Average Genetic Similarity

In comparison to changes in allelic counts which reflect changes in diversity at the gene level, changes in average genetic similarities indicate changes in diversity at the population level. The general overall trend was one of an increase in mean S_{ij} from the 1940s to the 2000s, although there was some oscillation between the 1950s and 1990s. Average S_{ij} values were the lowest for the 1940s (Table 3). The highest and second-highest average S_{ij} values (0.874 and 0.849) were for the 2000s and 1990s, respectively. The *t* test of Leonard et al. (1999) indicated the mean S_{ij} for the 2000s to be significantly higher than the mean S_{ij} for all other time periods except the 1990s ($P < 0.005$). The mean S_{ij} for the 1990s was found to be significantly higher than the mean S_{ij} for the initial gene pool, the 1940s, and the 1950s ($P < 0.03$).

UPGMA-Based Cultivar Grouping and Principle Co-ordinate Analysis

A UPGMA-based dendrogram was generated to gain insight on the relationships among the cultivars that were genotyped (Fig. 1). Cultivar groupings largely agreed with known pedigree information. Members of the group of cultivars released during the 1990s and 2000s appeared to be the most closely related as many of the cultivars from these two periods were clustered tightly together. Only 'NC606,' 'Speight NF3,' and 'CU748' resided outside of the grouping that included most cultivars of these two time periods. In general, cultivars from the two most recent time periods were more similar to each other than were cultivars of the initial gene pool and the 1940s. Cultivars of the latter two periods were much more loosely grouped.

Associations were further assessed using principle coordinate analysis. The first two axes from this analysis explained 34% of the total observed variation, with the first and second axes explaining 21 and 13% of the variation, respectively. Lines were drawn between extreme points of areas occupied by cultivars of three time periods (the initial gene pool, the 1940s, and the 2000s) following the procedure of Donini et al. (2000). This approach highlights the ranges of diversity for cultivars of each time period. Areas occupied by cultivars of the initial gene pool and the 1940s were large, and substantial overlap was observed between these two groups (Fig. 2). Breeding efforts appeared to increase the similarity of recently released material as cultivars of the 2000s occupied a much smaller area relative to cultivars of the first two periods. Cultivars of the first two time periods were widely dispersed on the scatterplot. Cultivars from the 2000s had much less dispersion and were shifted to the left on the plot. The plot showed very little overlap between cultivars of the 2000s and those from the initial gene pool and the 1940s.

DISCUSSION

Advanced cycle pedigree breeding has the desirable effect of reducing the frequency of, or eliminating, unfavorable alleles due to selection. Some alleles would also be expected to be lost due to genetic drift. A negative aspect of this breeding practice, however, is that genetic variability within modern populations can be substantially reduced. Modern cultivars are more elite, but also more narrow genetically.

Table 3. Average genetic similarity values (S_{ij}) for tobacco cultivars of eight different periods.

Time period	No. of cultivars	Average S_{ij}	Average S_{ij} SE
Initial gene pool	37	0.758	0.023
1940s	16	0.694	0.020
1950s	11	0.704	0.030
1960s	11	0.796	0.021
1970s	10	0.763	0.027
1980s	11	0.776	0.016
1990s	10	0.849	0.032
2000s	11	0.874	0.012

In this study, multiple methods incorporating analysis of microsatellite data were applied to investigate possible changes in genetic diversity in flue-cured tobacco germplasm. Results indicated that (i) only a fraction of the alleles (48%) present in the initial gene pool are represented in flue-cured tobacco cultivars of the 1990s and 2000s, (ii) genetic diversity has been gradually reduced at the gene level, and (iii) genetic diversity as measured at the population level is lowest for the 1990s and 2000s time periods. Similar studies have been conducted for several other crop species. Some reports have suggested loss of genetic diversity as a consequence of plant breeding to be negligible (Donini et al., 2000; Manifesto et al., 2001; Christiansen et al., 2002; Khan et al., 2005; Reif et al., 2005). Other studies have indicated changes in diversity to be of a qualitative nature (at the gene level), rather than at the population level (Russell et al., 2000; Lu and Bernardo, 2001; Roussel et al., 2004, 2005; Fu et al., 2005). Maccaferri et al. (2003) reported an increase in the level of diversity present in modern cultivars of durum wheat.

The reduction in genetic diversity for U.S. flue-cured tobacco is not surprising. Tobacco breeding began with a relatively narrow germplasm base. Murphy et al. (1987) found the average coefficient of parentage among a group of 131 flue-cured tobacco cultivars to be very high (0.41), and considerably higher than similar estimates for other self-pollinated crop species (Lewis and Nicholson, 2007). Results from the current study indicate that the core flue-cured germplasm base has become narrower with time. The high degree of genetic similarity among materials released during the 1990s and 2000s may be due, to a large extent, to many recent cultivars being largely developed from crosses involving 'K326,' a very popular high-yielding and high-quality cultivar released in the early 1980s. Narrow germplasm bases may be more evident for crop species such as barley (Martin et al., 1991) and tobacco (Murphy et al., 1987) where quality attributes are included amongst the most important breeding objectives. The situation for flue-cured tobacco may be an extreme example of the impact that stringent quality requirements and conservative breeding strategies can have on genetic diversity in germplasm pools. To be eligible for commercial release, experimental flue-cured tobacco cultivars must meet a set of rigid quality requirements outlined by the U.S. Regional Minimum Standards Program (Bowman, 1996). Almost all cultivars have been developed from crosses between elite materials, and "exotic" germplasm is thought to have entered into elite germplasm pools only in small doses through efforts to introgress simply inherited disease-resistance genes from wild *Nicotiana* relatives, or to incorporate oligogenic resistance to bacterial

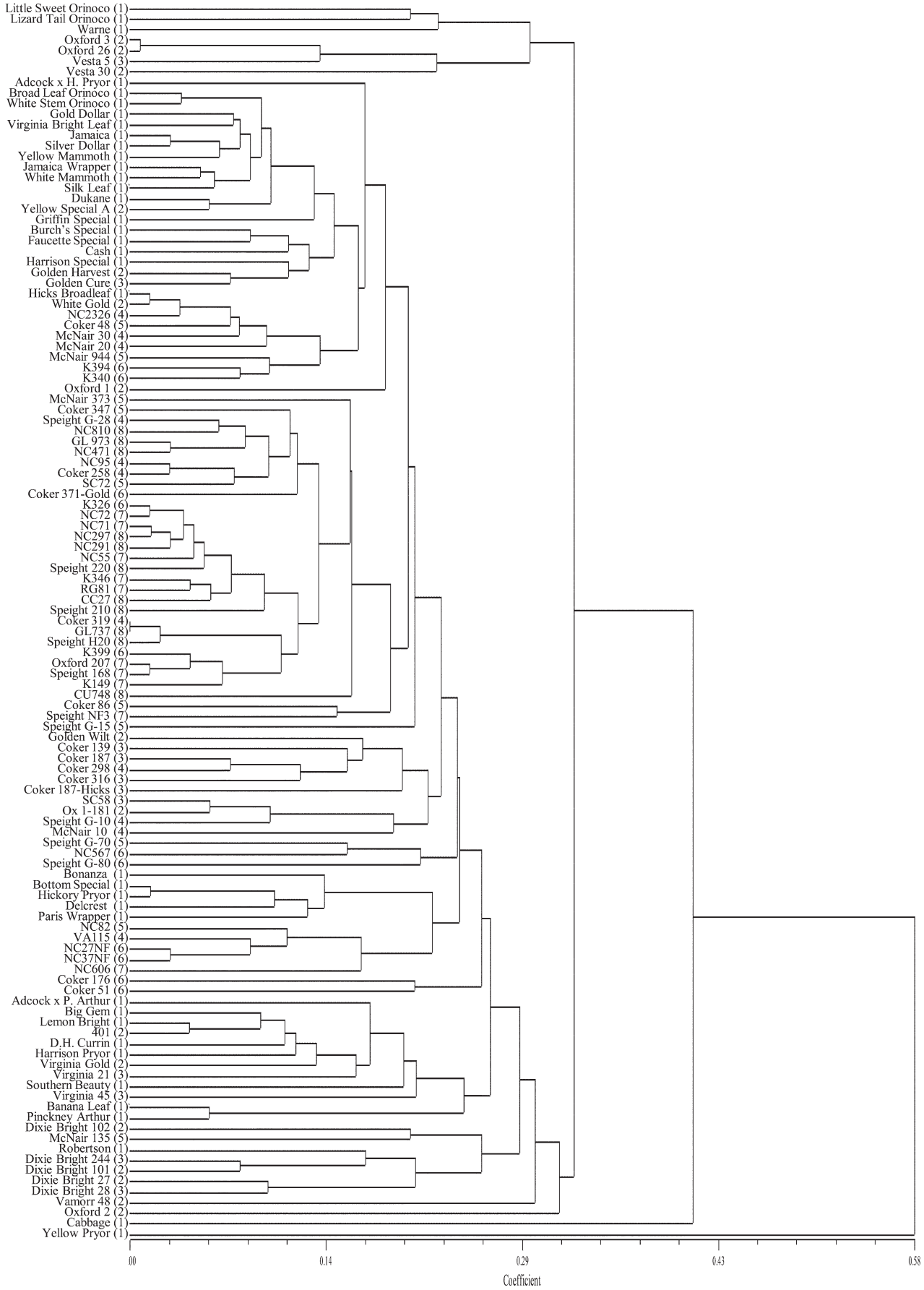

Figure 1. Microsatellite-based UPGMA dendrogram of 117 flue-cured tobacco cultivars. Numbers in parenthesis indicate breeding time period (1 = initial gene pool, 2 = 1940s, 3 = 1950s, 4 = 1960s, 5 = 1970s, 6 = 1980s, 7 = 1990s, and 8 = 2000s).

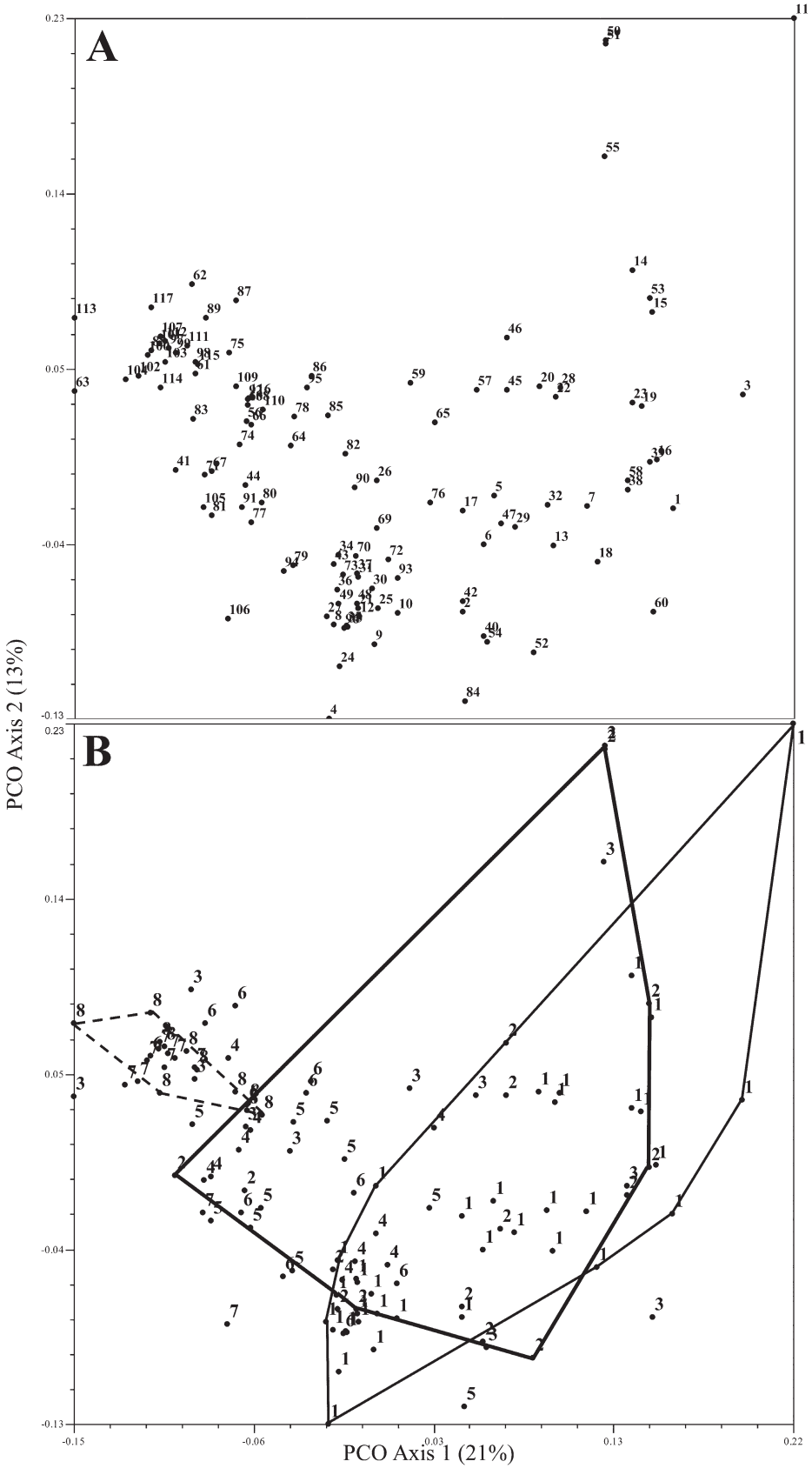

Figure 2. (A) Plot of the first two axes derived from principle co-ordinate (PCO) analysis of microsatellite data for 117 flue-cured tobacco cultivars. Numbers correspond to cultivar codes as listed in Table 1. (B) Plot for first two PCO axes, where numbers indicate breeding periods (1 = initial gene pool, 2 = 1940s, 3 = 1950s, 4 = 1960s, 5 = 1970s, 6 = 1980s, 7 = 1990s, and 8 = 2000s). Lines (regular, thickened, and dashed) join together variation extremities for the initial gene pool, 1940s, and 2000s, respectively.

wilt and black shank from *N. tabacum* lines 'Florida 301' and 'TI448A,' respectively (Bullock, 1943; Smith et al., 1945). In the current study, only 13 and 18 alleles were detected in the 1990s and 2000s, respectively, which were undetected in the initial gene pool. It may be worthwhile to investigate whether any of these markers are linked to genes affecting disease resistance.

Genetic variability is generally considered essential for gain from selection. Lower or reduced levels of genetic variability may restrict the potential for further genetic improvements for important agronomic characteristics. Genetic gain for yield had been steadily realized from the 1960s until the early 1980s (Bowman et al., 1984). This, in fact, can be cited as a success from continued selection within a very restricted germplasm pool. Very little genetic advancement has been made for yield in flue-cured tobacco since the 1980s, however. The marker-based results from the current study indicate an extremely high degree of genetic similarity among modern breeding materials. Increased attention may need to be paid to diversification of germplasm in flue-cured breeding populations if continued gains for yield are desired, or if breeders wish to maintain flexibility for dealing with future breeding challenges. Results indicate that a fair amount of allelic diversity exists within the U.S. flue-cured tobacco germplasm pool as a whole. Genetic relationships illustrated in Fig. 1 may provide valuable information for strategic germplasm choices in future tobacco breeding efforts.

Acknowledgments

The authors are grateful to Philip Morris USA and Philip Morris International for financial support of the NCSU tobacco genetics research program. The authors also thank Dr. Yong-Bi Fu of Agriculture and Agri-Food Canada for providing SAS code and helpful comments relative to this research project.

References

Bai, D., R. Reeleder, and J.E. Brandle. 1995. Identification of two RAPD markers tightly linked with the *Nicotiana debneyi* gene for resistance to black root rot of tobacco. Theor. Appl. Genet. 91:1184–1189.

Bernardo, R. 2002. Breeding for quantitative traits in plants. Stemma Press, Woodbury, MN.

Bindler, G., R. van der Hoeven, I. Gunduz, J. Plieske, M. Ganal, L. Rossi, F. Gadani, and P. Donini. 2007. A microsatellite marker based linkage map of tobacco. Theor. Appl. Genet. 114:341–349.

Bowman, D.T. 1996. History of the regional minimum standards program for the release of flue-cured tobacco varieties in the United States. Tob. Sci. 40:99–110.

Bowman, D.T., E.A. Wernsman, T.C. Corbin, and A.G. Tart. 1984. Contribution of genetics and production technology to long-term yield and quality gains in flue-cured tobacco. Tob. Sci. 28:30–35.

Bullock, J.F. 1943. Strains of flue-cured tobacco resistant to black shank. USDA Circ. 682. U.S. Gov. Print. Office, Washington, DC.

Chakraborty, R., and L. Jin. 1993. Estimation of genetic distance and coefficient of gene diversity from single-probe multilocus DNA fingerprinting data. Mol. Biol. Evol. 11(1):120–127.

Chaplin, J.F., T.W. Graham, and Z.T. Ford. 1962. Agronomic and chemical characteristics of certain old-line varieties of flue-cured tobacco. Tob. Sci. 6:16–20.

Christiansen, M.J., S.B. Anderson, and R. Ortiz. 2002. Diversity changes in an intensively bred wheat germplasm during the 20th century. Mol. Breed. 9:1–11.

Clayton, E.E. 1958. The genetics and breeding progress in tobacco during the last 50 years. Agron. J. 50:352–356.

Dice, L.R. 1945. Measures of the amount of ecologic association between species. Ecology 26:297–302.

Donini, P., J.R. Law, R.M.D. Koebner, J.C. Reeves, and R.J. Cooke. 2000. Temporal trends in the diversity of UK wheat. Theor. Appl. Genet. 100:912–917.

East, E.M., and D.F. Jones. 1921. Round tip tobacco: A plant "made to order." J. Hered. 12:50–56.

Fu, Y.-B., W.G. Peterson, K.W. Richards, D. Somers, R.M. DePauw, and J.M. Clarke. 2005. Allelic reduction and genetic shift in the Canadian hard red spring wheat germplasm released from 1845 to 2004. Theor. Appl. Genet. 110:1505–1516.

Fu, Y.-B., W.G. Peterson, G. Scoles, B. Rossnagel, D.J. Schoen, and K.W. Richards. 2003. Allelic diversity changes in 96 Canadian oat cultivars released from 1886 to 2001. Crop Sci. 43:1989–1995.

Fu, Y.-B., W.G. Peterson, J.-K. Yu, L. Gao, J. Jia, and K.W. Richards. 2006. Impact of plant breeding on genetic diversity of the Canadian hard red spring wheat germplasm as revealed by EST-derived SSR markers. Theor. Appl. Genet. 112:1239–1247.

Garner, W.W., H. Allard, and E.E. Clayton. 1936. Superior germplasm in tobacco. p. 785–830. *In* 1936 Yearbook of Agriculture. USDA, Washington, DC.

Johnson, E.S., M.F. Wolff, and E.A. Wernsman. 2002. Marker-assisted selection for resistance to black shank disease in tobacco. Plant Dis. 86:1303–1309.

Khan, I.A., F.S. Awan, A. Ahman, Y.B. Fu, and A. Iqbal. 2005. Genetic diversity of Pakistan wheat germplasm as revealed by RAPD markers. Genet. Resour. Crop Evol. 52:239–244.

Le Clerc, V., F. Bazante, C. Baril, J. Guiard, and D. Zhang. 2005. Assessing temporal changes in genetic diversity of maize varieties using microsatellite markers. Theor. Appl. Genet. 110:294–302.

Leonard, A.C., S.E. Franson, V.S. Hertzberg, M.K. Smith, and G.P. Toth. 1999. Hypothesis testing with the similarity index. Mol. Ecol. 8:2105–2114.

Lewis, R.S. 2005. Transfer of resistance to potato virus Y (PVY) from *Nicotiana africana* to *Nicotiana tabacum*: Possible influence of tissue culture on the rate of introgression. Theor. Appl. Genet. 110:678–687.

Lewis, R.S., S.R. Milla, and S.P. Kernodle. 2007. Analysis of an introgressed *Nicotiana tomentosa* genomic regions affecting leaf number and correlated traits in *Nicotiana tabacum*. Theor. Appl. Genet. 114:841–854.

Lewis, R.S., S.R. Milla, and J.S. Levin. 2005. Molecular and genetic characterization of *Nicotiana glutinosa* L. chromosome segments in tobacco mosaic virus-resistant tobacco accessions. Crop Sci. 45:2355–2362.

Lewis, R.S., and J.S. Nicholson. 2007. Aspects of the evolution of *Nicotiana tabacum* L. and the status of the United States *Nicotiana* germplasm collection. Genet. Resour. Crop Evol. 54:727–740.

Liu, K., and S.V. Muse. 2005. PowerMarker: An integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129.

Lu, H., and R. Bernardo. 2001. Molecular marker diversity among current and historical maize inbreds. Theor. Appl. Genet. 103:613–617.

Maccaferri, M., M.C. Sanguinetti, P. Donini, and R. Tuberosa. 2003. Microsatellite analysis reveals a progressive widening of the genetic base in the elite durum wheat germplasm. Theor. Appl. Genet. 107:783–797.

Manifesto, M.M., A.R. Schlatter, H.E. Hopp, E.Y. Suarez, and J. Dubcovsky. 2001. Quantitative evaluation of genetic diversity in wheat germplasm using molecular markers. Crop Sci. 41:682–690.

Martin, J.M., T.K. Blake, and E.A. Hockett. 1991. Diversity among North American spring barley cultivars based on coefficients of parentages. Crop Sci. 31:1131–1137.

Milla, S.R., J.S. Levin, R.S. Lewis, and R.C. Rufty. 2005. RAPD and SCAR markers linked to an introgressed gene conditioning resistance to *Peronospora tabacina* D.B. Adam. in tobacco. Crop Sci. 45:2346–2354.

Moon, H.S., and J.S. Nicholson. 2007. Identification of AFLP markers linked to *Tomato spotted wilt virus* resistance in tobacco. Crop Sci. 47:1887–1894.

Murphy, J.P., T.S. Cox, R.C. Rufty, and D.M. Rodgers. 1987. A representation of the pedigree relationships among flue-cured tobacco cultivars. Tob. Sci. 31:70–75.

Nishi, T., T. Tajima, S. Noguchi, H. Ajisaka, and H. Nedishi. 2003. Identification of DNA markers of tobacco linked to bacterial wilt resistance. Theor. Appl. Genet. 106:765–770.

Reif, J.C., P. Zhang, S. Dreisigacher, M.L. Warburton, M. van Ginkel, D. Hoisington, M. Bohn, and A.E. Melchinger. 2005. Wheat genetic diversity trends during domestication and breeding. Theor. Appl. Genet. 110:859–864.

Ren, N., and M.P. Timko. 2001. AFLP analysis of genetic polymorphism and evolutionary relationships among cultivated and wild *Nicotiana* species. Genome 44:559–571.

Rohlf, F.J. 2000. NTSYS-pc: Numerical taxonomy and multivariate analysis system, version 2.2. Exeter Software, Setauket, NY.

Rossi, L., G. Bindler, H. Pijnenburg, P.G. Isaac, I. Giraud-Henry, M. Mahe, C. Orvain, and F. Gadani. 2001. Potential of molecular marker analysis for variety identification in processed tobacco. Plant Var. Seeds 14:89–101.

Roussel, V., J. Koenig, M. Bechert, and F. Balfourier. 2004. Molecular diversity in French bread wheat accessions related to temporal trends and breeding programmes. Theor. Appl. Genet. 108:920–930.

Roussel, V., L. Leisova, F. Exbrayat, Z. Stehno, and F. Balfourier. 2005. SSR allelic diversity changes in 480 European bread wheat varieties released from 1840 to 2000. Theor. Appl. Genet. 111:162–170.

Russell, J.R., R.P. Ellis, W.T.B. Thomas, R. Waugh, J. Provan, A. Booth, J. Fuller, P. Lawrence, G. Young, and W. Powell. 2000. A retrospective analysis of spring barley germplasm development from 'foundation genotypes' to currently successful cultivars. Mol. Breed. 6:553–568.

Shamel, A.D., and W.W. Cobey. 1907. Tobacco breeding. USDA Bureau Plant Ind. Bull. No. 96. U.S. Gov. Print. Office, Washington, DC.

Smith, T.E., E.E. Clayton, and E.G. Moss. 1945. Flue-cured tobacco resistant to bacterial (Granville) wilt. USDA Circ. No. 727. U.S. Gov. Print. Office, Washington, DC.

Sneath, P.H.A., and R.R. Sokal. 1973. Numerical taxonomy. Freeman, San Francisco, CA.

Wernsman, E.A., and R.C. Rufty. 1987. Tobacco. p. 669–698. *In* W.R. Fehr (ed.) Principles of cultivar development. Vol. 2. Crop species. Macmillian, New York.

Yap, I.V., and R.J. Nelson. 1996. WinBoot: A program for performing bootstrap analysis of binary data to determine the confidence limits of UPGMA-based dendrograms. International Rice Research Institute, Manila, Philippines.

Yi, Y.-H., R.C. Rufty, and E.A. Wernsman. 1998. Identification of RAPD markers linked the wildfire resistance gene of tobacco using bulked segregant analysis. Tob. Sci. 42:52–57.